

<https://helda.helsinki.fi>

Statistical ecology comes of age

Gimenez, Olivier

2014-12-01

Gimenez , O , Buckland , S T , Morgan , B J T , Bez , N , Bertrand , S , Choquet , R , Dray , S , Etienne , M-P , Fewster , R , Gosselin , F , Merigot , B , Monestiez , P , Morales , J M , Mortier , F , Munoz , F , Ovaskainen , O , Pavoine , S , Pradel , R , Schurr , F M , Thomas , L , Thuiller , W , Trenkel , V , de Valpine , P & Rexstad , E 2014 , ' Statistical ecology comes of age ' , Biology Letters , vol. 10 , no. 12 , 20140698 . <https://doi.org/10.1098/rsbl.2014.0698>

<http://hdl.handle.net/10138/225309>

<https://doi.org/10.1098/rsbl.2014.0698>

cc_by

publishedVersion

Downloaded from Helda, University of Helsinki institutional repository.

This is an electronic reprint of the original article.

This reprint may differ from the original in pagination and typographic detail.

Please cite the original version.



Cite this article: Gimenez O *et al.* 2014
Statistical ecology comes of age. *Biol. Lett.* **10**:
20140698.

<http://dx.doi.org/10.1098/rsbl.2014.0698>

Received: 29 August 2014

Accepted: 4 December 2014

Subject Areas:

ecology

Keywords:

citizen science, hidden Markov model,
hierarchical model, movement ecology,
software package, spatially explicit
capture–recapture, species distribution
modelling, state–space model

Author for correspondence:

Olivier Gimenez

e-mail: olivier.gimenez@cefe.cnrs.fr

Electronic supplementary material is available
at <http://dx.doi.org/10.1098/rsbl.2014.0698> or
via <http://rsbl.royalsocietypublishing.org>.

Statistical ecology comes of age

Olivier Gimenez¹, Stephen T. Buckland², Byron J. T. Morgan³, Nicolas Bez⁴,
Sophie Bertrand⁴, Rémi Choquet¹, Stéphane Dray⁵, Marie-Pierre Etienne⁶,
Rachel Fewster⁷, Frédéric Gosselin⁸, Bastien Mérigot⁹, Pascal Monestiez¹⁰,
Juan M. Morales¹¹, Frédéric Mortier¹², François Munoz¹³, Otso Ovaskainen¹⁴,
Sandrine Pavoine^{15,16}, Roger Pradel¹, Frank M. Schurr¹⁷, Len Thomas²,
Wilfried Thuiller¹⁸, Verena Trenkel¹⁹, Perry de Valpine²⁰ and Eric Rexstad²

¹CEFE UMR 5175, CNRS, Université de Montpellier, Université Paul-Valéry Montpellier, EPHE,
1919 Route de Mende, 34293 Montpellier Cedex 5, France

²Centre for Research into Ecological and Environmental Modelling, University of St Andrews,
St Andrews KY16 9LZ, UK

³School of Mathematics, Statistics and Actuarial Science, University of Kent, Canterbury, Kent CT2 7NF, UK

⁴IRD, UMR EME 212, Sète, France

⁵Université de Lyon, F-69000, Lyon; Université Lyon 1; CNRS, UMR5558, Laboratoire de 18 Biométrie et Biologie
Evolutive, F-69622, Villeurbanne, France

⁶AgroParisTech, UMR MIA 518, Paris, France

⁷Department of Statistics, University of Auckland, Private Bag 92019, Auckland, New Zealand

⁸Irstea, UR EFNO, Centre de Nogent-sur-Vernisson, 45290 Nogent-sur-Vernisson, France

⁹Université Montpellier 2, UMR EME 212, Sète, France

¹⁰INRA, BioSP, Avignon, France

¹¹Laboratorio Ecotono, CRUB, INIBIOMA-CONICET, Bariloche, Argentina

¹²UPR Bsef, CIRAD, Montpellier, France

¹³UM2, UMR AMAP, Bd de la Lironde, TA A-51/PS2, 34398 Montpellier Cedex 5, France

¹⁴Department of Biosciences, University of Helsinki, Helsinki, Finland

¹⁵UMR 7204 CNRS UPMC, Centre for Ecology and Conservation Sciences, Muséum National d'Histoire Naturelle,
55-61 rue Buffon, 75005 Paris, France

¹⁶Mathematical Ecology Research Group, Department of Zoology, University of Oxford, Oxford OX1 3PS, UK

¹⁷Institute of Landscape and Plant Ecology, University of Hohenheim, 70593 Stuttgart, Germany

¹⁸Laboratoire d'Ecologie Alpine, UMR CNRS 5553, Université Joseph Fourier, Grenoble I, BP 53,
38041 Grenoble Cedex 9, France

¹⁹Ifremer, Rue de l'île d'Yeu, BP 21105, 44311 Nantes Cedex 3, France

²⁰Environmental Science, Policy and Management, University of California, Berkeley, CA 94720, USA

JMM, 0000-0001-7269-7490

The desire to predict the consequences of global environmental change has been the driver towards more realistic models embracing the variability and uncertainties inherent in ecology. Statistical ecology has gelled over the past decade as a discipline that moves away from describing patterns towards modelling the ecological processes that generate these patterns. Following the fourth International Statistical Ecology Conference (1–4 July 2014) in Montpellier, France, we analyse current trends in statistical ecology. Important advances in the analysis of individual movement, and in the modelling of population dynamics and species distributions, are made possible by the increasing use of hierarchical and hidden process models. Exciting research perspectives include the development of methods to interpret citizen science data and of efficient, flexible computational algorithms for model fitting. Statistical ecology has come of age: it now provides a general and mathematically rigorous framework linking ecological theory and empirical data.

1. Introduction

Variability is challenging ecology, from genes to individuals, species or ecosystems: quantifying and explaining biological variation is an ever-important goal. Variability arises from both ecological processes and sampling, requiring the modelling of uncertainty, the very nature of statistics [1,2].

Statistics has long permeated the field of ecology through the contributions of eminent scientists such as Fisher, Haldane and Leslie. However, we detect a recent rise in statistical awareness, manifested in various ways. First, research centres especially devoted to statistical ecology have been created in the USA (Statistical and Applied Mathematical Sciences Institute) and the UK (National Centre for Statistical Ecology). There are also institutes focused on synthesis (e.g. the National Center for Ecological Analysis and Synthesis and the National Institute for Mathematical and Biological Synthesis, both in the USA). Second, new journals dedicated to methodological advances (not only statistical) have been created and are now having considerable impact (notably *Molecular Ecology Resources* and *Methods in Ecology and Evolution*). Third, there are more specialized conferences that provide the opportunity for statisticians to interact with ecologists for mutual benefit. The reasons for this recent rise of statistical ecology are manifold and include the societal demand for scientists to address pressing issues such as global change and the current biodiversity crisis, the need to analyse the massive datasets and the novel data types generated by new technologies, and the popularization of methods through free statistical packages and the increase in computing power. We view the rise of statistical ecology as a sign that ecological and statistical modelling are coming together with the common goal of understanding complex processes in a formal inferential framework for better predictive capabilities. We acknowledge that not all ecologists agree that ecology lends itself to theorization and prediction [3] or that process-based methods necessarily have higher predictive ability than phenomenological models [4,5]. However, past disappointments may simply be due to inappropriate and coarse modelling. If so, progress in both ecological theory and statistical ecology and a better integration of the two should enhance our understanding and our ability to predict ecological phenomena. In the following, we highlight recent trends in statistical ecology and provide perspectives for the future development of this discipline (see also [6]).

We analysed the contents of the abstracts of four International Statistical Ecology Conferences (ISECs) held biannually between 2008 and 2014 to provide a picture of recent trends in statistical ecology (electronic supplementary material, Appendix S1). The quantitative results of this analysis show a temporal shift across the different ISECs, from studies focusing on sampling design issues towards predictive studies that aim to integrate the modelling of processes with the analysis of ecological patterns. These results are further synthesized below.

2. Questions being addressed

(a) Assessing species distribution

Species distribution models (SDMs) are now common tools to investigate the main drivers of species range and to forecast potential impacts of environmental changes on biodiversity. Important innovations include the use of point processes to fit SDMs to presence-only data and the mathematical equivalence

of MAXENT (a common SDM tool) to generalized linear models in this context [7]. SDMs are also being extended to several species to improve the model parametrization for rare species and to enable the estimation of co-occurrence patterns. Last, the development of hierarchical occupancy models, with their ability to handle spatial dependence and imperfect detection, paves the way for better modelling of the underlying sources of uncertainty [8].

(b) Measuring biodiversity

Biodiversity is multifaceted, involving aspects of species richness, functions, traits and phylogeny. Consequently, the choice of relevant diversity indices is challenging, especially when analysing aspects of functional or phylogenetic diversity and when evaluating the dissimilarities among locations (quadrats, sites or regions). Moreover, the potential factors driving the dynamics of biodiversity (e.g. competition and environmental filters) need to be disentangled.

(c) Investigating population dynamics

In the ISECs, estimation of population size has been a major focus, notably through refinements of capture–recapture (CR) methods. There has been an increase in non-invasive methods that use natural identifying characteristics of animals (camera or acoustic traps, genetic markers), with treatment of misidentification error. In parallel, spatially explicit models have been developed to fully exploit the spatial information in CR data [9,10].

(d) Understanding animal movements

Movement ecology has shifted from phenomenological models of observable patterns to mechanistic models characterizing the underlying processes. In particular, the use of state–space models that account explicitly for the observation process has now become standard [11], and hierarchical models have been developed to model individual movements as functions of behavioural states, past experiences and environmental heterogeneity [12]. While earlier work relied on discrete-time correlated random walks, the use of continuous-time models and the integration of other types of data (e.g. species interactions and population dynamics) are increasing.

(e) Interpreting citizen science data

Data from citizen science programmes represent an opportunity to sample large regions and inform long-term monitoring studies. Difficulties arise with recent programmes based on web- and smartphone-based technologies that are characterized by the free participation of many laypersons, loose sampling protocols and heterogeneities in the spatio-temporal distribution of observations. These potential sources of bias may be accounted for by the joint modelling of the ecological and observation processes through, for example, hidden process models [13].

3. Material and methods

(a) Hidden process modelling

Ecologists have broadly adopted hierarchical, state–space and hidden Markov models to deal with the way in which individuals

and populations distribute in space and change over time [14]. This reflects a move away from modelling spatio-temporal patterns *per se* and towards modelling the ecological processes that generate those patterns. The timescale of interest might be short, such as for animal behaviour, medium, such as migration and demographic processes, or long, such as for changes in species ranges, composition and biodiversity, or for evolutionary processes. By modelling the underlying processes while accounting for observation error and model uncertainty, we seek to gain in predictive ability and hence in the effectiveness of management actions, whether we are managing a commercial fishery, conserving a threatened population, assessing the impact on biodiversity of habitat loss, predicting response of populations to disturbance or evaluating the effects of climate change on communities.

(b) Coexistence of frequentist and Bayesian frameworks

Bayesian methods are now widely used, largely because they can more easily accommodate realistic ecological models. However, two notable trends are emerging: an increasing interest in critically evaluating the performance of Bayesian methods from a frequentist perspective [15], and the increasing practicality of frequentist tools for hierarchical models previously only amenable to Bayesian methods (e.g. [16]).

(c) Dynamic models

Current research in population dynamics addresses the limits of statistical inference and predictions for nonlinear dynamics (e.g. [17]). Beyond the population, dynamic statistical models are now applied at larger spatial and organizational scales to describe the dynamics of species ranges, communities and ecosystem processes (e.g. [18]). A common feature of these recent statistical models is that they describe how large-scale dynamics arise from underlying principles of demography and/or ecophysiology, aiming to base inference and prediction on processes rather than correlations.

(d) Integrated modelling

Another trend is the popularization of integrated modelling—i.e. combining different datasets in a single, coherent analysis [19]—to address a wide variety of ecological questions. Current developments deal with the issues of goodness-of-fit testing, model selection, integration of recent developments in demography (e.g. integral projection models) and testing the assumption that data from different sources can be considered independent. From an ecological viewpoint, integrated modelling now scales from populations up to communities [20].

4. Implementation

(a) Computational algorithms

The development of efficient and flexible computational algorithms for complex models and big datasets ([integrated nested] Laplace approximations, Hamiltonian Monte Carlo and standard Markov chain Monte Carlo algorithms) requires tremendous research efforts, as does their implementation in software packages (e.g. R-INLA (<http://www.r-inla.org/>), AD Model Builder (<http://admb-project.org/>), Laplaces Demon (<http://www.bayesian-inference.com/software>), Stan (<http://mc-stan.org/>), Nimble (<http://r-nimble.org/>), OpenBUGS (<http://www.openbugs.net/w/FrontPage>), JAGS (<http://mcmc-jags.sourceforge.net/>), PyMC (<http://pymc-devs.github.io/pymc/>), MCMCglmm (<http://cran.r-project.org/web/packages/MCMCglmm/index.html>)).

When a complete likelihood cannot be easily calculated, methods for estimation based only on simulations and summary statistics (Synthetic likelihood [21]; Approximate Bayesian Computation [22]) are also receiving attention.

(b) Software development and evaluation

There is a tension between devoting time to developing new methodology and enabling other researchers to implement it. Although it is easy to self-publish an R package or a graphical user interface (GUI), a culture shift is needed towards more thorough testing and verification of published software. We welcome the initiative of ecological journals to publish software papers, which ensures that publicly available software is peer-reviewed, and endows software development efforts with much-needed professional recognition.

5. Advice to statistical ecologists

(a) Avoiding statistical machismo¹

Given methodological developments and increasing computing power, there is a great temptation to increase model complexity. In some cases, this is helpful: previously restrictive assumptions about the observation process can be relaxed; previously intractable ecological mechanisms can be expressed as mathematical models and incorporated in estimation. In other cases, however, increasing complication can lead to less robust inference or ecologically insignificant improvements, which nevertheless waste practitioners' time and direct their energies away from less glamorous topics such as improved data collection; there is also often an increased chance of mistakes in implementation. There is a clear need for an evaluation strategy of new, often complex statistical methods to determine the scope of beneficial application for ecology [23]. *Beneficial* means that for a given ecological question and dataset, applying the new or modified method provides clearer results and avoids drawing flawed conclusions. Comprehensive model evaluation must include consideration of sample design, covariate selection, goodness-of-fit and parameter redundancy diagnostics.

(b) Going one step further

Many ecological applications are motivated by scientific support for conservation or management decisions. Statistical decision theory has much to offer, both directly in terms of helping rational decision-making, and also in optimizing future data-collection efforts.

6. Conclusion

The dialogue between statisticians and ecologists has intensified over recent decades, and ISECs have contributed to this dialogue. We encourage even more mixing between statisticians and ecologists, by exhorting the former to go to the field to gain a sound understanding of the data for relevant modelling [24], and the latter to embrace courses in mathematics that underpin the reliable application of statistical methods [25].

In summary, the statistical approaches developed for ecology are maturing towards a statistically rigorous, explanatory

and possibly predictive framework for linking theory, data and applications. Exciting research directions are ahead of us that we hope will help to address pressing issues in the context of global change.

Acknowledgements. We thank the scientific and local organizing committees who largely contributed to the success of the ISECs. This is a contribution of the GDR 3645 'Statistical Ecology'.

Authors' contributions. All authors participated in the study, drafted the manuscript and gave final approval for publication.

Conflict of interests. The authors declare no competing interests.

Endnote

¹See <http://dynamicceology.wordpress.com/2012/09/11/statistical-machismo/>.

References

- Davidian M, Louis TA. 2012 Why statistics? *Science* **336**, 12. (doi:10.1126/science.1218685)
- Spiegelhalter D. 2014 The future lies in uncertainty. *Science* **345**, 264–265. (doi:10.1126/science.1251122)
- Cooper GJ. 2003 *The science of the struggle for existence: on the foundations of ecology*. Cambridge, UK: Cambridge University Press.
- Peters H. 1991 *A critique for ecology*. Cambridge, UK: Cambridge University Press.
- Breiman L. 2001 Statistical modeling: the two cultures. *Stat. Sci.* **16**, 199–231. (doi:10.1214/ss/1009213726)
- King R. 2014 Statistical ecology. *Annu. Rev. Stat. Appl.* **1**, 401–426. (doi:10.1146/annurev-statistics-022513-115633)
- Renner IW, Warton DI. 2013 Equivalence of MAXENT and Poisson point process models for species distribution modeling in ecology. *Biometrics* **69**, 274–281. (doi:10.1111/j.1541-0420.2012.01824.x)
- MacKenzie DI, Nichols JD, Royle JA, Pollock KH, Bailey LL, Hines JE. 2006 *Occupancy estimation and modeling: inferring patterns and dynamics of species occurrence*. Oxford, UK: Academic Press.
- Borchers DL, Stevenson BC, Kidney D, Thomas L, Marques TA. 2014 A unifying model for capture–recapture and distance sampling surveys of wildlife populations. *J. Am. Stat. Assoc.* (doi:10.1080/01621459.2014.893884)
- Royle JA, Chandler RB, Sollmann R, Gardner B. 2014 *Spatial capture–recapture*. New York, NY: Academic Press.
- Patterson TA, Thomas L, Wilcow C, Ovaskainen O, Matthiopoulos J. 2008 State–space models of individual animal movement. *Trends Ecol. Evol.* **23**, 87–94. (doi:10.1016/j.tree.2007.10.009)
- McClintock BT, King R, Thomas L, Matthiopoulos J, McConnell BJ, Morales J. 2012 A general discrete-time modeling framework for animal movement using multistate random walks. *Ecol. Monogr.* **82**, 335–349. (doi:10.1890/11-0326.1)
- Pagel J, Anderson BJ, O'Hara RB, Cramer W, Fox R, Jeltsch F, Roy DB, Thomas CD, Schurr FM. 2014 Quantifying range-wide variation in population trends from local abundance surveys and widespread opportunistic occurrence records. *Methods Ecol. Evol.* **5**, 751–760. (doi:10.1111/2041-210X.12221)
- Clark JS. 2007 *Models for ecological data*. Princeton, NJ: Princeton University Press.
- Little R. 2011 Calibrated Bayes, for statistics in general, and missing data in particular. *Stat. Sci.* **26**, 162–174. (doi:10.1214/10-STS318)
- Lele SR, Dennis B, Lutscher F. 2007 Data cloning: easy maximum likelihood estimation for complex ecological models using Bayesian Markov chain Monte Carlo methods. *Ecol. Lett.* **10**, 551–563. (doi:10.1111/j.1461-0248.2007.01047.x)
- Hartig F, Dormann CF. 2013 Does model-free forecasting really outperform the true model? *Proc. Natl Acad. Sci. USA* **110**, E3975. (doi:10.1073/pnas.1308603110)
- Clark JS, Agarwal P, Bell DM, Flikkema PG, Gelfand A, Nguyen X, Ward E, Yang J. 2011 Inferential ecosystem models, from network data to prediction. *Ecol. Appl.* **21**, 1523–1536. (doi:10.1890/09-1212.1)
- Newman KB, Buckland ST, Morgan BJT, King R, Borchers DL, Cole DJ, Besbeas P, Gimenez O, Thomas L. 2014 *Modelling population dynamics—model formulation, fitting and assessment using state–space methods*. New York, NY: Springer.
- Péron G, Koons DN. 2012 Integrated modeling of communities: parasitism, competition, and demographic synchrony in sympatric ducks. *Ecology* **93**, 2456–2464. (doi:10.1890/11-1881.1)
- Wood SN. 2010 Statistical inference for noisy nonlinear ecological dynamic systems. *Nature* **466**, 1102–1104. (doi:10.1038/nature09319)
- Csilléry K, Blum M, Gaggiotti O, François O. 2010 Approximate Bayesian computation (ABC) in practice. *Trends Ecol. Evol.* **25**, 410–418. (doi:10.1016/j.tree.2010.04.001)
- Hodges J. 2010 Are exercises like this a good use of anybody's time? *Ecology* **91**, 3496–3500. (doi:10.1890/10-0052.1)
- Gimenez O *et al.* 2013 How can quantitative ecology be attractive to young scientists? Balancing computer/desk work with field work. *Anim. Conserv.* **16**, 134–136. (doi:10.1111/j.1469-1795.2012.00597.x)
- Barraquand F, Ezard THG, Jørgensen PS, Zimmerman N, Chamberlain S, Salguero-Gómez R, Curran TJ, Poisot T. 2014 Lack of quantitative training among early-career ecologists: a survey of the problem and potential solutions. *PeerJ.* **2**, e285. (doi:10.7717/peerj.285)